

Облачные технологии для развития территориально распределенной вычислительно-аналитической геологической среды

В. С. ЕРЁМЕНКО^{1,*}, В. В. НАУМОВА¹, А. А. ЗАГУМЕННОВ², С. В. БУЛОВ¹

¹Государственный геологический музей им. В. И. Вернадского РАН, Москва, 125009, Россия

²Институт автоматизации и процессов управления ДВО РАН, Владивосток, 690041, Россия

*Контактный автор: Ерёмченко Виталий Сергеевич, e-mail: vitaer@gmail.com

Поступила 29 июля 2020 г., доработана 7 октября 2020 г., принята в печать 17 ноября 2020 г.

Описываются подходы к организации единого рабочего пространства исследователя для обработки геологической информации. Предложены подходы по взаимодействию с внешними территориально распределенными сервисами обработки и анализа геологических данных на основе WPS-платформы и по организации интегральной платформы для доступа к интерактивным облачным сервисам обработки геологических данных. Приводятся примеры облачных сервисов и платформ и сравниваются их основные характеристики. Описывается программная реализация предложенных подходов в рамках разрабатываемой информационно-аналитической системы GeologyScience.ru. Приводится общая архитектура системы, и кратко описывается каждый из ее модулей. Описываются внешние сервисы, интегрированные в разрабатываемую платформу.

Ключевые слова: облачный сервис, web-сервис, Web Processing Service, программная платформа.

Цитирование: Ерёмченко В.С., Наумова В.В., Загуменнов А.А., Булов С.В. Облачные технологии для развития территориально распределенной вычислительно-аналитической геологической среды. Вычислительные технологии. 2021; 26(1):86–98. DOI:10.25743/ICT.2021.26.1.007.

Введение

Обработка данных является важной частью исследовательских работ в любой предметной области. С развитием информационных технологий менялись подходы к организации такой обработки. Для решения задачи анализа больших объемов данных или при использовании ресурсоемких методов появилась потребность в ускорении программных вычислений. Для этого использовались как параллельные вычисления с применением суперкомпьютеров, так и распределенные вычисления, использовавшие большое количество вычислительных устройств, объединенных в единую вычислительную систему, такие как GRID-системы. Однако доступ к таким системам довольно ограничен. Появление облачных сервисов позволило пользователям получить простой доступ к интересующим их вычислительным ресурсам.

Облачные сервисы обеспечивают доступ к компьютерному оборудованию, аппаратным ресурсам, дисковой памяти, базам данных и операционным системам в удаленном режиме. В широком смысле к облачным сервисам можно отнести услуги, доступ к которым предоставляется через Интернет, в интерфейсе браузера либо при помощи устанавливаемой клиентской программы.

Согласно работе Louise Moser [1], облачные сервисы можно разделить на три модели обслуживания пользователей по типу предоставляемых услуг:

- программное обеспечение как сервис (Software as a Service, SaaS);
- платформа как сервис (Platform as a Service, PaaS);
- инфраструктура как сервис (Infrastructure as a Service, IaaS).

Поставщики сервисов из группы SaaS размещают у себя программное обеспечение (ПО) и предоставляют к нему доступ с использованием облачных технологий. Такой подход освобождает пользователя от необходимости устанавливать ПО самостоятельно, а также перекладывает ответственность за обслуживание и обновление ПО на поставщика сервиса.

Поставщики сервисов из группы PaaS предлагают пользователям вычислительные платформы, которые могут включать в себя операционные системы, программные среды, web-серверы и базы данных. Конечные пользователи могут разрабатывать и запускать программное обеспечение непосредственно на облачной платформе без необходимости покупки и установки собственной локальной платформы.

Поставщики сервисов из группы IaaS предоставляют пользователям физические компьютеры, хранилища, сети, а также средства управления этими ресурсами. Пользователи могут разворачивать на таких узлах свои собственные образы операционных систем, устанавливать прикладные приложения и поддерживать собственное программное обеспечение.

Разработчики программных решений по обработке данных активно внедряют в процесс обработки облачные вычисления (группа SaaS). Программный продукт по обработке данных размещается в виде web-сервиса, доступного через сеть Интернет.

Использование облачных сервисов для решения научных задач позволяет исследователям получить доступ к различным специализированным программным инструментам без необходимости покупки, загрузки, установки отдельных программных приложений и дальнейшего их обслуживания.

В сфере науки и образования использование облачных инфраструктур избавляет от необходимости инвестировать колоссальные средства в создание собственных дата-центров, в закупку серверного оборудования и лицензированного ПО, в зарплату системных администраторов и компьютерных техников.

1. Обзор облачных сервисов для решения геологических задач

Среди облачных сервисов первой группы, которые можно использовать для анализа и обработки научных данных, выделяют несколько основных направлений:

- Облачные сервисы, предназначенные для хранения информации. Основные характеристики сервисов представлены в табл. 1.
- Облачные сервисы для предварительной обработки данных: текстовые, табличные процессоры и сервисы для создания презентаций (табл. 2).
- Сервисы для тематического анализа данных: пространственных, спутниковых и т. п. (табл. 3).

- Сервисы визуализации данных (табл. 4).
- Службные сервисы: онлайн-конвертеры документов, изображений, аудио- и видеофайлов и т. п. (табл. 5).

Т а б л и ц а 1. Основные характеристики облачных сервисов хранения данных
Table 1. Basic parameters of cloud storage services

Характеристика	Dropbox	Яндекс.Диск	Google Drive	Облако @Mail.ru	OneDrive
Объем бесплатного дискового пространства, Гб	2–48	10–15	15	25	15
Максимальный размер загружаемых файлов, Гб	–	10	10	2–32	2–4
Возможность разграничения прав доступа	–	+	+	–	+
Контроль изменений	+	–	+	–	+
Создание общих папок	+	+	+	+	+
Загрузка файлов по почте	+	+	+	+	+
Редактирование файлов через интерфейс	–	+	+	–	+
Просмотр файлов через web-интерфейс	+	+	+	+	+

Т а б л и ц а 2. Характеристики облачных сервисов редактирования документов и презентаций
Table 2. Characteristics of cloud services for editing documents and presentations

Сервис	OnlyOffice	Google Docs	MS Office Online	Zoho Docs
Форматы документов	DOC, DOCX, DOTX, ODT, OTT, RTF, TXT, HTML, ODS, OTS, XLS, XLSX XLTX, CSV, PPT, PPTX, POTX, ODP, OTP	DOC, DOCX, ODT, TXT, PPT, PPTX, XPS, RTF, CSV, TXT, HTML	DOC, DOCX, XLS, XLSX, PPT, PPTX, TXT, CSV	DOC, ODF, ODT, SXW, RTF, HTML, TXT, PDF, LaTeX
Поддержка кириллицы в редакторе	+	+	+	+
Контроль получателей	+	+	+	+
Общий доступ к документам	+	+	+	+
Авторизация для пользователей при совместной работе	Обычная регистрация	Google ID, персональное приглашение	Windows Live ID, Facebook ID	Google ID, Facebook ID, Yahoo ID, обычная регистрация

Т а б л и ц а 3. Характеристики облачных сервисов тематического анализа пространственных и спутниковых данных

Table 3. Characteristics of cloud services for thematic analysis of spatial and satellite data

Характеристика	ArcGIS Online	Google Earth Engine
Интерактивная работа со слоями	+	+
Загрузка пользовательских данных	+	+
Наличие собственного каталога пространственных данных	+	+
Возможность создания собственных процедур обработки	+	+ (JavaScript, Python)

Т а б л и ц а 4. Характеристики облачных сервисов визуализации данных

Table 4. Characteristics of data visualization cloud services

Сервис	Поддержка источников данных
Google Data Studio	Google-таблицы, Cloud SQL, MySQL, PostgreSQL, CSV, XLSX и т. д.
Power BI	Базы данных различных форматов, данные Google и Microsoft, файлы табличных данных и т. д.
Tableau	XML, MS Excel, MySQL, SQL, Google BigQuery, Microsoft Azure и т. д.
ChartBlocks	Базы данных различных форматов, электронные таблицы и т. д.
Plotly	Excel, базы данных MySQL, Redshift, некоторые web-сервисы и т. д.
Infogram	Excel, MySQL, PostgreSQL, Amazon Redcliff, Microsoft SQL Server и т. д.

Т а б л и ц а 5. Характеристики сервисов конвертации документов и медиафайлов

Table 5. Characteristics of services for converting documents and media files

Сервис	Тип поддерживаемых данных
CloudConvert	Архивы, документы, изображения, аудио- и видеофайлы, книги, презентации, шрифты, таблицы и т. д.
Free File Converter	Архивы, документы, изображения, аудио- и видеофайлы, книги, презентации, шрифты, таблицы и т. д.
Swift Converter	Документы, изображения, аудио- и видеофайлы, 3D-графика
FileConverto	Изображения, аудио- и видеофайлы

2. Информационно-вычислительные среды

Для обработки и анализа научных данных в сетевой среде создаются различные тематические территориально распределенные вычислительные среды для научных целей.

В работе L. Candela и др. [2] приводится общий обзор виртуальных исследовательских сред, выделяются общие и отличительные особенности различных подходов к построению таких сред и разбираются проблемы, которые необходимо решать в данной области.

А.М. Федотовым, В.Б. Барахниным и др. [3] предложена концепция распределенной информационно-аналитической среды для исследования экологических систем. В статье описывается модель виртуальной среды, определяются категории данных, объекты среды и приводится пример схемы среды с описанием используемых технологий.

Е.П. Гордовым, В.Н. Крупчатниковым и др. [4] представлен проект по созданию тематической виртуальной исследовательской среды для анализа, оценки и прогнози-

рования воздействия глобального изменения климата. Проект среды разрабатывается с целью обеспечения свободного доступа к различным ресурсам данных и службам обработки через web-браузер.

И.В. Бычковым, Г.М. Ружниковым и др. [5, 6] разработана и успешно функционирует среда WPS-сервисов обработки геоданных, которая поддерживает вызов сервисов обработки, построенных с использованием интерфейса OGC WPS (Web Processing Service). В среде реализована возможность построения цепочек обработки с использованием языка JavaScript для формирования сценария обработки.

3. Вычислительно-аналитическая геологическая среда ГГМ РАН

С 2017 г. в Государственном геологическом музее им. В. И. Вернадского РАН (ГГМ РАН) ведутся работы по созданию и адаптации методов и технологий обработки и анализа территориально распределенной разнотипной геологической информации и сервисов ее обработки. На основе созданных подходов, разработанных методов и технологий реализуется информационно-аналитическая среда для поддержки и сопровождения научных исследований в геологии, осуществляющая интеграцию территориально распределенной геологической информации с использованием специализированных служб ее анализа и обработки (<http://geologyscience.ru>) (рис. 1) [7]. Разработанная программная платформа управления тематическими сервисами обработки и анализа является частью

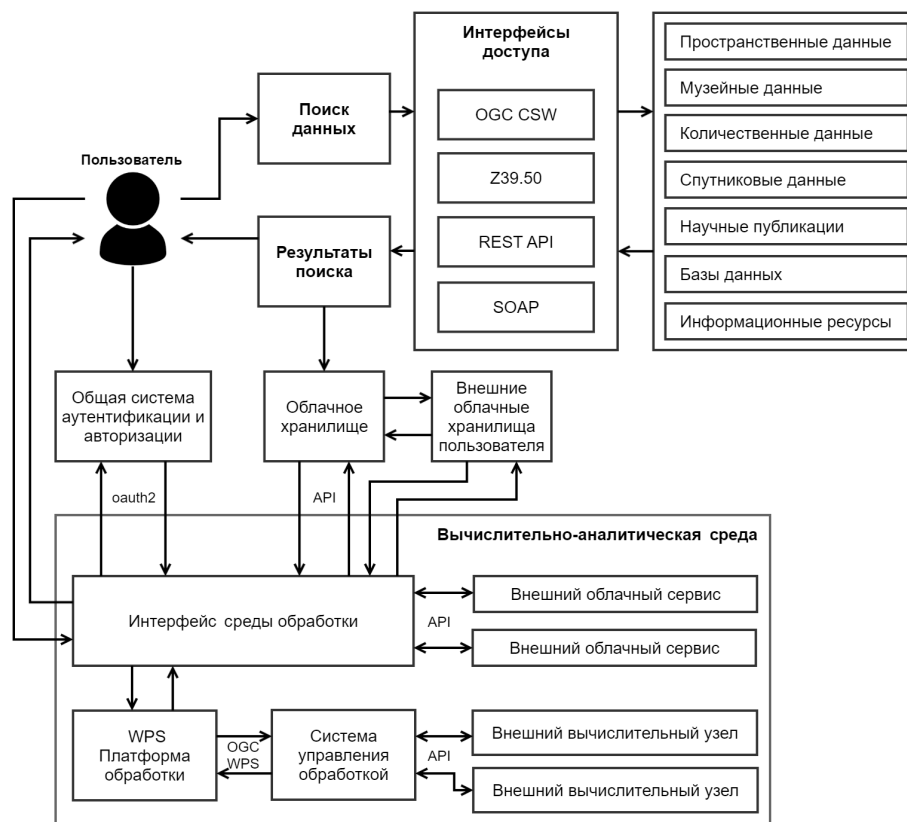


Рис. 1. Общая функциональная схема информационно-аналитической среды GeologyScience.ru
Fig. 1. General functional diagram of the Information and Analytical Environment GeologyScience.ru

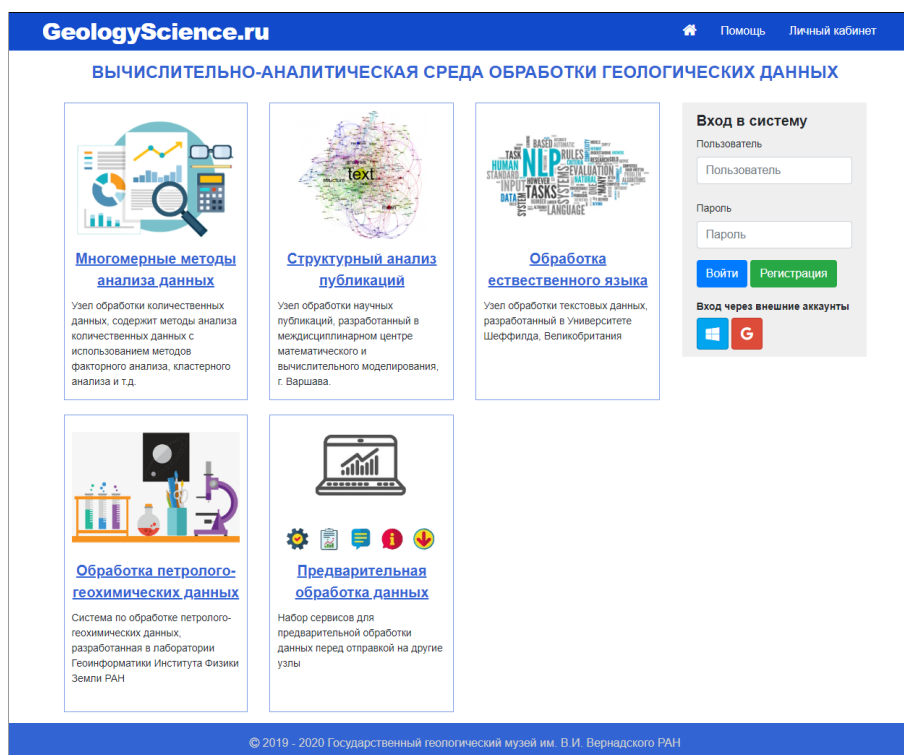


Рис. 2. Web-интерфейс главной страницы информационно-аналитической среды GeologyScience.ru

Fig. 2. Web interface of the Information and Analytical Environment GeologyScience.ru

информационно-аналитической среды, она обеспечит пользователям доступ к хранилищам современных наукоемких алгоритмов и вычислительных ресурсов, необходимых для оперативной обработки больших массивов геологических данных.

Цель вычислительного блока информационно-аналитической среды (далее — вычислительно-аналитической геологической среды) — предоставление исследователям доступа к территориально распределенным сервисам обработки и анализа геологической информации (<https://service.geologyscience.ru>) (рис. 2) [8]. В первой версии вычислительно-аналитической геологической среды для организации единого механизма взаимодействия с web-сервисами используется интерфейс OGC Web Processing Service. Для каждого внешнего web-сервиса создан соответствующий ему WPS-процесс, реализующий интерфейс взаимодействия с внешним вычислительным узлом.

В вычислительной среде реализован доступ к следующим вычислительным узлам и платформам:

- *Обработка табличных данных.* Вычислительный узел “Многомерные методы анализа данных” разработан в ГГМ РАН. Он позволяет выполнять обработку табличных данных различными методами анализа данных с настройкой их параметров и визуализацией результатов. Вычислительный узел включает в себя такие группы методов, как статистический анализ, регрессионный анализ, факторный анализ, кластеризация, машинное обучение, методы визуализации и др. Вычисления производятся в среде Python с использованием известных пакетов обработки данных: Scikit-learn, Pandas, Matplotlib и др. Обработка поступающих запросов производится фреймворком Flask посредством REST API с использованием очереди заданий, реализованной на основе NoSQL базы данных Redis. Такая архитектура позволяет разделить обработку запросов и тяжелые вычисления больших

объемов данных, что обеспечивает отказоустойчивость и масштабируемость узла. В дальнейшем планируется расширение количества методов обработки данных с привлечением специализированных пакетов обработки для решения геологических задач.

- *Обработка петролого-геохимических данных.* В Институте физики земли им. О. Ю. Шмидта РАН разработана интерактивная база методов обработки петролого-геохимических данных [9], которая предоставляет сервисы: построения спайдерграмм, гистограмм и классификационных диаграмм, идентификации минералов по их химическому составу, интерпретации состава минерала и разложения на минералы и т. д. Интерфейс взаимодействия с сервисами построен на основе REST-архитектуры.
- *Структурный анализ публикаций.* В междисциплинарном центре математического и вычислительного моделирования (Университет Варшавы, Польша) разработан сервис для извлечения метаданных из научных публикаций [10]. Метаданные включают в себя авторов, аффилиацию, абстракт, ключевые слова, название журнала, объем, год выпуска, разобранные библиографические ссылки, структуру разделов документа, заголовки разделов и абзацы. Интерфейс взаимодействия с сервисами построен на основе REST-архитектуры.
- *Обработка естественного языка.* В Университете Шеффилда в рамках проекта GATE (General Architecture for Text Engineering) создан ряд сервисов по обработке текстовых данных для различных языков [11]. Для обработки текстовых данных на русском языке предоставляются сервисы по определению частей речи, а также выделению именованных сущностей, таких как имена и фамилии, названия организаций, географические названия, даты, денежные единицы и т. д. Интерфейс взаимодействия с сервисами построен на основе REST-архитектуры.

Для отслеживания состояния территориально распределенных компонентов среды разработан модуль мониторинга, позволяющий с установленной периодичностью проверять доступность удаленных узлов и работоспособность используемых сервисов по указанным протоколам доступа. В качестве источника данных для мониторинга используется модуль каталога внешних сервисов обработки и анализа геологической информации [12].

Первая версия программной среды реализована в формате web-приложения на языке программирования Java. Для размещения используется контейнер web-приложений с открытым исходным кодом Apache Tomcat (<http://tomcat.apache.org>). В качестве платформы реализации WPS-сервиса используется программный пакет с открытым исходным кодом GeoServer (<http://geoserver.org>). Система мониторинга также реализована в виде самостоятельного web-приложения на языке программирования Java.

Развитие информационных технологий позволило производителям программных продуктов перенести некоторые настольные приложения по обработке и анализу данных в облачную инфраструктуру. Использование таких облачных сервисов для исследователей значительно расширяет набор инструментов для анализа данных.

Тенденции развития информационных технологий требуют разработки подхода, позволяющего интегрировать такие облачные сервисы в едином пользовательском пространстве. Создание единого протокола доступа к интерактивным облачным сервисам представляет определенные трудности в силу разнородности сервисов. Вторая версия разрабатываемой среды предполагает новый подход к интеграции интерактивных облачных сервисов на основе единого пользовательского пространства данных.

4. Интерактивные облачные сервисы во второй версии вычислительно-аналитической геологической среды: Excel, ArcGIS Online, Google Earth Engine

Одним из наиболее популярных программных продуктов обработки *табличных данных* в интерактивном режиме является Excel из пакета Microsoft Office. Этот программный продукт содержит ряд инструментов для редактирования данных, построения различных диаграмм, использования встроенных процедур анализа и создания собственных. Компания Microsoft разработала бесплатную облачную версию продукта Excel (<https://www.office.com>). Зарегистрированный пользователь может использовать полноценную web-версию для данных, расположенных в облачном хранилище Microsoft One Drive.

Для анализа *пространственных данных* наиболее подходящим инструментом являются геоинформационные системы, позволяющие в интерактивном режиме взаимодействовать с объектами на карте, применяя к отдельным слоям и объектам доступные средства анализа. Наиболее подходящим для геологии облачным решением является ArcGIS Online, разрабатываемый компанией ESRI. Portal for ArcGIS — это инфраструктура ArcGIS Online, функционирующая в защищенной ИТ-среде или в частном облаке организации (под контролем сетевого экрана или в полностью изолированной локальной сети). Портал позволяет создавать карты, каталогизировать и анализировать пространственные данные с помощью удобного, интуитивно понятного интерфейса. Создание и публикация интерактивных карт и приложений может выполняться на любом устройстве при наличии браузера и доступа к Интернету. Публикация данных в защищенном облаке ESRI в виде кэшированных или динамических сервисов соответствует всем современным стандартам защиты информации с сохранением всех авторских прав.

Для анализа *спутниковых данных* одним из лидеров среди облачных сервисов является платформа Google Earth Engine (<https://earthengine.google.com>), которая позволяет пользователю загружать собственные данные или использовать данные из каталога Earth Engine для дальнейшей обработки в интерактивном режиме. В каталоге содержатся продукты обработки данных радиометра Modis (спутники Aqua, Terra), спутников Sentinel-1A, Sentinel-1B, Sentinel-2A, Sentinel-2B, Landsat 8 и др. Earth Engine содержит ряд предустановленных алгоритмов анализа, а также инструменты для их создания, редактирования и запуска с использованием языков программирования Javascript и Python. Для работы пользователю необходимо наличие аккаунта Google. Для анализа и обработки можно использовать данные из облачного хранилища Google.

5. Интеграция интерактивных облачных сервисов в вычислительную среду

Каждый из перечисленных облачных сервисов использует собственные уникальные протоколы взаимодействия, существенно затрудняя задачу интеграции сервиса в разрабатываемую вычислительную среду на основе общего протокола доступа.

Некоторые облачные сервисы требуют наличия данных в их собственном хранилище. Например, при использовании Excel из MS Office Online данные должны находиться в пользовательском хранилище One Drive. Использование Earth Engine позволяет также использовать данные из облачного хранилища Google на аккаунте пользователя. Таким

образом, чтобы обеспечить “бесшовный” переход между облачными сервисами необходимо разработать набор процедур для публикации выбранных пользователем данных в соответствующем пользовательском облачном хранилище. Создание подобной процедуры стало возможным при использовании технологии web-приложений, позволяющей запросить у пользователя разрешение на доступ к определенным возможностям пользовательского аккаунта различных поставщиков облачных сервисов.

Данная технология поддерживается Microsoft, Google, Yandex, ESRI и т. д. Перед использованием соответствующего сервиса пользователю предлагается данные для анализа загрузить в его персональное хранилище. Для этого он проходит авторизацию на сайте поставщика сервиса, после чего приложение запрашивает у пользователя разрешение на скачивание и публикацию данных в его хранилище. При выборе облачного сервиса другого поставщика пользователь имеет возможность перемещения данных для обработки в хранилище данного поставщика.

Таким образом, нами предложены подход и технологическое решение для организации единого пространства данных для различных поставщиков облачных сервисов.

6. Описание модулей разрабатываемой платформы и принципов их взаимодействия

Платформа по обработке и анализу геологических данных представляет собой программный продукт на языке программирования Java в формате web-приложения с использованием фреймворка Spring Boot. Платформа реализована на основе принципов слабосвязанной архитектуры. Для взаимодействия с пользователем реализован web-интерфейс с применением JavaScript фреймворка AngularJS. Также разработан программный интерфейс с использованием архитектурного стиля REST для взаимодействия с платформой в формате система — система.

Для выполнения поставленных задач разработан ряд авторских модулей.

WSIM — модуль взаимодействия с внешними web-сервисами. Основная функция модуля — обеспечение единого интерфейса доступа к внешним сервисам обработки и анализа, работающим по принципу запрос — ответ. В качестве интерфейса доступа используется OGC Web Processing Service 1.0. Каждый используемый вычислительно-аналитической средой метод публикуется в виде отдельного процесса внутри WPS-сервиса. Для реализации WPS-сервиса используется программный пакет с открытым исходным кодом GeoServer. GeoServer содержит встроенные компоненты для реализации очереди заданий, разграничения доступа к различным процессам и возможности как синхронного, так и асинхронного выполнения процесса. Для реализации модуля используются программные библиотеки GeoTools и 52 North WPS. Взаимодействие платформы с данным модулем осуществляется через WPS-интерфейс.

MICS — модуль взаимодействия с общим облачным хранилищем пользователя. Основная функция модуля — обеспечение интерфейса доступа к облачному хранилищу пользователя, используемому в информационно-аналитической системе GeologyScience.ru. Пользователь имеет возможность поместить данные, выбранные в информационной системе GeologyScience.ru, в облачное хранилище для дальнейшего их использования в других подсистемах, в том числе для обработки и анализа. Результаты обработки также могут быть помещены пользователем в облачное хранилище для их последующего использования в других сервисах. Взаимодействие с платформой осуществляется через REST-интерфейс.

GAM — модуль общей авторизации. Основная функция модуля — обеспечение единой точки аутентификации и авторизации для пользователей всех блоков информационно-аналитической среды GeologyScience.ru. Модуль обращается к сервису авторизации на базе подхода Single Sign-On (SSO) с использованием протокола OAuth. Для реализации взаимодействия модуля с сервисом аутентификации и авторизации используется фреймворк Spring Security.

CCM — модуль взаимодействия с внешними облачными хранилищами пользователей. Основная функция модуля — организация единого пространства данных на основе имеющихся у пользователей аккаунтов поставщиков услуг облачного хранения для дальнейшего использования данных при работе с разнородными интерактивными облачными сервисами. Поставщики облачных сервисов часто требуют предварительной загрузки данных для обработки в их собственные облачные хранилища. Модуль позволит исследователю при переходе от одного сервиса к другому в автоматическом режиме перемещать данные в требуемое хранилище. Данный модуль является основным инструментом, позволяющим интегрировать разнородные интерактивные сервисы в рамках разрабатываемой платформы. Взаимодействие с платформой осуществляется через REST-интерфейс.

SCM — модуль каталога внешних сервисов. Основные функции модуля — обеспечение пользователя информацией об имеющихся в среде сервисах с возможностью получения информации о каждом зарегистрированном сервисе, включая общее описание, данные о поставщике сервиса и техническую информацию для организации взаимодействия. Информация из данного модуля используется в модуле мониторинга для отслеживания состояния каждого внешнего сервиса, зарегистрированного в вычислительно-аналитической среде. Взаимодействие с платформой осуществляется с использованием REST-интерфейса.

SSMM — модуль мониторинга состояния системы и сервисов. Основная функция модуля — отслеживание статуса каждого внешнего поставщика сервиса, зарегистрированного в среде, а также всех блоков системы GeologyScience.ru. Мониторинг состояния сервисов производится по нескольким параметрам, включая доступность узла, на котором размещен сервис, доступность сервиса по указанному протоколу и проверку сервиса на изменения путем отправки тестовых запросов к сервису. Взаимодействие с платформой осуществляется через REST-интерфейс.

В результате проведенных работ авторами создана цифровая платформа по обработке и анализу геологических данных DPAР (рис. 3).

Таким образом, серверная технология DPAР обеспечивает безопасное подключение к среде совместной работы, которая превращает пользовательские разрозненные данные в безопасный для анализа фонд информации. Масштабируемая технология создания каталогов и управления большими наборами данных позволяет администрировать одновременный доступ с разными пользовательскими настройками и допусками безопасности. Платформа позволяет пользователям эффективно каталогизировать, визуализировать большие пространственные данные и управлять ими, что поможет минимизировать риск и принимать более обоснованные решения, увеличивающие возможность успеха обнаружения месторождений. Использование цифровой платформы по обработке и анализу геологических данных позволит в дальнейшем расширять возможности разработчиков информационно-аналитических сред для поддержки и сопровождения научных исследований.

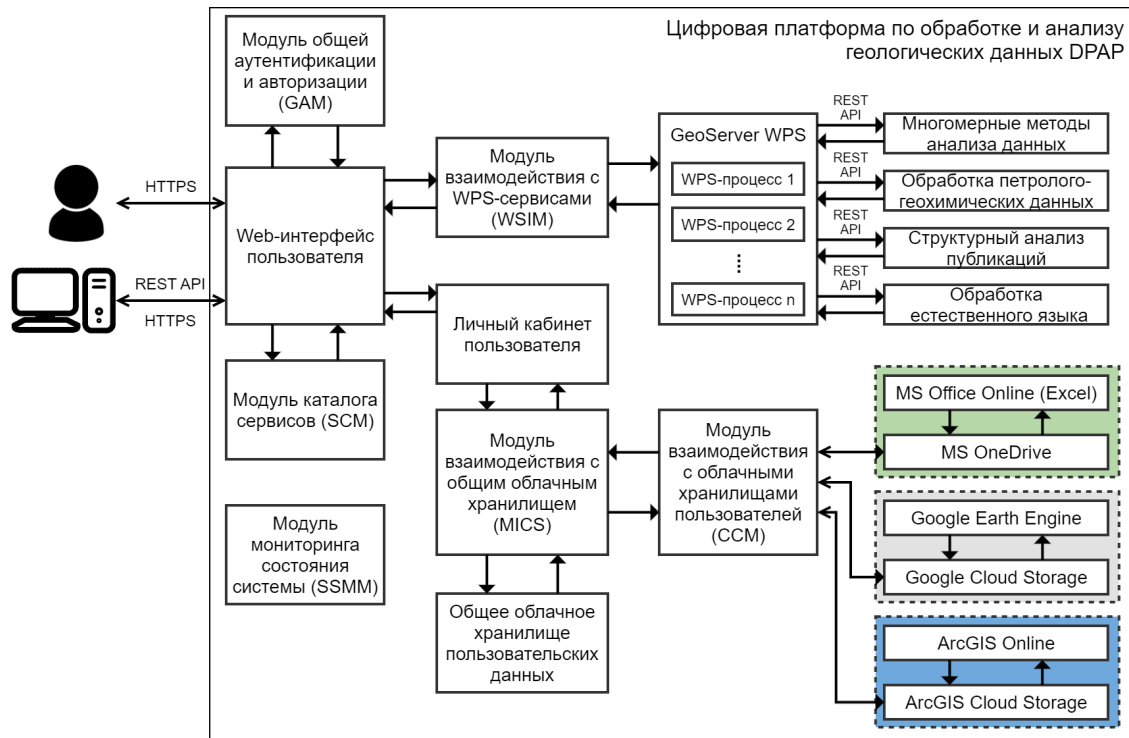


Рис. 3. Общая структурная схема DPAP

Fig. 3. General DPAP block diagram

Благодарности. Работы выполняются в рамках государственного задания ГГМ РАН по теме № 0140-2019-0005 “Разработка информационной среды интеграции данных естественнонаучных музеев и сервисов их обработки для наук о Земле”.

Список литературы

- [1] Moser L., Thuraisingham B., Zhang J. Services in the Cloud. *IEEE Transactions on Services Computing*. 2015; 8(2):172–174. DOI:10.1109/TSC.2015.2410351. Available at: <https://ieeexplore.ieee.org/document/7055320>.
- [2] Candela L., Castelli D., Pagano P. Virtual research environments: An overview and a research agenda. *Data Science Journal*. 2013; (12):GRDI75–GRDI81. DOI:10.2481/dsj.GRDI-013.
- [3] Фёдоров А.М., Баряхнин В.Б., Гуськов А.Е., Молородов Ю.И. Распределенная информационно-аналитическая среда для исследований экологических систем. Вычислительные технологии. Спецвыпуск. Избранные доклады Международной конференции по измерениям, моделированию и информационным системам для изучения окружающей среды: ENVIROMIS-2006 (Томск, Россия). Ч. 1. 2006; (11):113–125.
- [4] Gordov E.P., Krupchatnikov V.N., Okladnikov I.G., Fazliev A.Z. Thematic virtual research environment for analysis, evaluation and prediction of global climate change impacts on the regional environment. *Proc. SPIE. 22nd International Symposium on Atmospheric and Ocean Optics: Atmospheric Physics*, 100356J. 2016; (10035):1394–1403. DOI:10.1117/12.2249118. Available at: <https://doi.org/10.1117/12.2249118>.
- [5] Бычков И.В., Ружников Г.М., Фёдоров Р.К., Шумилов А.С. Компоненты среды WPS-сервисов обработки геоданных. *Вестник НГУ. Сер.: Информационные технологии*. 2014; 12(3):16–24.

- [6] **Бычков И.В., Ружников Г.М., Фёдоров Р.К., Шумилов А.С.** Система планирования и выполнения композиций веб-сервисов в гетерогенной динамической среде. Вычислительные технологии. 2016. 21(6):18–35.
- [7] **Наумова В.В., Платонов К.А., Еременко В.С., Патук М.И., Дьяков С.Е.** Информационно-аналитическая среда для поддержки научных исследований в геологии: Текущее состояние и перспективы развития. Труды XVII Международной конференции “Распределенные информационно-вычислительные ресурсы. Цифровые двойники и большие данные. (DICR-2019)”. Новосибирск; 2019: 139–147. DOI:10.25743/ICT.2019.70.61.021.
- [8] **Eremenko V.S., Naumova V.V., Platonov K.A., Dyakov S.E., Eremenko A.S.** The main components of a distributed computational and analytical environment for the scientific study of geological systems. Russian Journal of Earth Sciences. 2018; (18):ES6003. DOI:10.2205/2018ES00063.
- [9] **Иванов С.Д.** Интерактивный реестр геосенсоров на основе веб-приложения. Компьютерные исследования и моделирование. 2016; 8(4):621–632.
- [10] **Tkaczyk D., Szostek P., Fedoryszak M., Dendek P., Bolikowski L.** CERMINE: Automatic extraction of structured metadata from scientific literature. International Journal on Document Analysis and Recognition. 2015; 18(4):317–335. DOI:10.1007/s10032-015-0249-8.
- [11] **Maynard D., Bontcheva K., Augenstein I.** Synthesis lectures on the Semantic Web: Theory and technology. Natural Language Processing for the Semantic Web. 2016: 194. DOI:10.2200/S00741ED1V01Y201611WBE015.
- [12] **Ерёмченко В.С., Наумова В.В.** Система каталогизации и мониторинга территориально распределенных вычислительных узлов в среде WPS сервисов для решения геологических задач. Вестник НГУ. Сер.: Информационные технологии. 2019; 17(2):39–48.

Cloud technologies for development of geographically distributed computational and analytical Geological environment

EREMENKO VITALIY S.^{1,*}, NAUMOVA VERA V.¹, ZAGUMENNOV ALEXEY A.², BULOV STANISLAV V.¹

¹Vernadsky State Geological Museum RAS, Moscow, 125009, Russia

²Institute of Automation and Control Processes FEB RAS, Vladivostok, 690041, Russia

*Corresponding author: Eremenko Vitaliy S., e-mail: vitaer@gmail.com

Received July 29, 2020, revised October 7, 2020, accepted November 17, 2020

Abstract

The purpose of the study is to create an approach for organization of unified workspace for geological data research and processing. The study proposes approaches for organization of interaction with external geographically distributed processing and analytical services of geological data based on WPS platforms. An approach for integral platform organization providing access to interactive cloud processing of geological data is proposed. It connects the users with modern methods of data processing and analysis. A brief description of existing cloud services and platforms and comparison of their basic features is presented. Various groups of cloud services are described

depending on the type of services provided to the user. Software implementation of approaches proposed within developing informational and analytical system GeologyScience.ru is addressed. The general architecture of the system and author's modules developed including technological solutions used are described. A brief review of the services integrated into the platform being developed now is given. The use of the developed platform for processing and analyzing of geological data allows further expanding of capabilities of information-analytical environment developers to support scientific research.

Keywords: cloud service, web service, Web Processing Service, software platform.

Citation: Eremenko V.S., Naumova V.V., Zagumennov A.A., Bulov S.V. Cloud technologies for development of geographically distributed computational and analytical Geological environment. Computational Technologies. 2021; 26(1):86–98. DOI:10.25743/ICT.2021.26.1.007. (In Russ.)

Acknowledgements. The study is supported by the Government contract No. 0140-2019-0005 with SGM RAS “Development of an information environment for integrating data from natural science museums and their processing services for Earth sciences”.

References

1. Moser L., Thuraishingham B., Zhang J. Services in the Cloud. IEEE Transactions on Services Computing. 2015; 8(2):172–174. DOI:10.1109/TSC.2015.2410351. Available at: <https://ieeexplore.ieee.org/document/7055320>.
2. Candela L., Castelli D., Pagano P. Virtual research environments: An overview and a research agenda. Data Science Journal. 2013; 12:GRDI75–GRDI81. DOI:10.2481/dsj.GRDI-013.
3. Fedotov A.M., Barakhnin V.B., Guskov A.E., Molorodov Yu.I. Distributed information-analytical environment for ecological ecosystems investigations. Computational Technologies. Izbrannye doklady Mezhdunarodnoy konferentsii po izmereniyam, modelirovaniyu i informatsionnym sistemam dlya izucheniya okruzhayushchey sredy: ENVIROMIS-2006 (Tomsk, Rossiya). Chast' 1. 2006; (11):113–125. (In Russ.)
4. Gordov E.P., Krupchatnikov V.N., Okladnikov I.G., Fazliev A.Z. Thematic virtual research environment for analysis, evaluation and prediction of global climate change impacts on the regional environment. Proc. SPIE. 22nd International Symposium on Atmospheric and Ocean Optics: Atmospheric Physics, 100356J. 2016; (10035):1394–1403. DOI:10.1117/12.2249118. Available at: <https://doi.org/10.1117/12.2249118>.
5. Bychkov I.V., Ruzhnikov G.M., Fjodorov R.K., Shumilov A.S. Components of WPS environment for geoprocessing. Vestnik NGU. Seriya: Informatsionnye Tekhnologii. 2014; 12(3):16–24. URL: <https://lib.nsu.ru/xmlui/handle/nsu/6863>. (In Russ.)
6. Bychkov I.V., Ruzhnikov G.M., Fedorov R.K., Shumilov A.S. System for dynamic execution of composition services in the heterogeneous environment. Computational Technologies. 2016; 21(6):18–35. (In Russ.)
7. Naumova V.V., Platonov K.A., Eremenko V.S., Patuk M.I., Dyakov S.E. Information and analytical environment for supporting scientific research in geology: current state and development prospects. Proceedings of the XVII International conference “Distributed information and computing resources. (DICR-2019)”. Novosibirsk; 2019: 139–147. DOI:10.25743/ICT.2019.70.61.021. (In Russ.)
8. Eremenko V.S., Naumova V.V., Platonov K.A., Dyakov S.E., Eremenko A.S. The main components of a distributed computational and analytical environment for the scientific study of geological systems. Russian Journal of Earth Sciences. 2018; (18):ES6003. DOI:10.2205/2018ES00063.
9. Ivanov S.D. Web-based interactive registry of the geosensors. Computer Research and Modeling. 2016; 8(4):621–632. (In Russ.)
10. Tkaczyk D., Szostek P., Fedoryszak M., Dendek P., Bolikowski L. CERMINE: Automatic extraction of structured metadata from scientific literature. International Journal on Document Analysis and Recognition. 2015; 18(4):317–335. DOI:10.1007/s10032-015-0249-8.
11. Maynard D., Bontcheva K., Augenstein I. Synthesis lectures on the Semantic Web: Theory and technology. Natural Language Processing for the Semantic Web. 2016: 194. DOI:10.2200/S00741ED1V01Y201611WBE015.
12. Eremenko V.S., Naumova V.V. The system of cataloguing and monitoring of geographically distributed computing nodes in WPS services environment for geological problems solving. Vestnik NGU. Seriya: Informatsionnye Tekhnologii. 2019; 17(2):39–48. (In Russ.)